Learning Enhancement Team

University of East Anglia
STUDENT SUPPORT SERVICE

Steps into Statistics

Measurements of Central Tendency

This guide introduces two different measurements of central tendency: the mean and the median. It shows how to calculate them and gives guidance for using them appropriately.

Introduction

In statistics it is extremely useful to have a measure of the value of the middle of a data set. This can be thought of as a **typical** or **average** value but is more correctly called the **measurement of central tendency**. The word average is ambiguous as there are different ways of calculating and describing the centre of a data set. Measurements of central tendency are part of the group of fundamental **descriptive statistics**.

Many descriptions of research which include statistics will be interpreted in terms of an "average" and you should be comfortable making informed decisions regarding such statements. For example, the "average" yearly student debt for England is £4729. This figure should not suggest to you that every student in England has debts of £4729. One interpretation could be that if the total debt of students in England was evenly distributed between them, each would owe £4729. Another would be that if all the student debts were written out in order from smallest to largest then £4729 would be directly in the middle.

Measurements of central tendency are useful as they can summarise some aspects of information regarding a large amount of data into a single value. However they should not be used as the only descriptive statistic for a data set as they give no information about the variability of the data within a data set. Measurements of central tendency are combined with measurements of spread to give a better description of a data set (see study guides*: Measurements of Spread I: Range and Inter-Quartile Range* and *II: Variance and Standard Deviation*).

You may have an inherent understanding of what is meant by an "average" but, because there are different levels of data in statistics (see study guide*: Levels of Data*), there are different ways of calculating 'the middle' of a data set. Two measures of the middle of a data set are the **mean** and the **median**. The centre of certain data sets can be described by either of these and each may be a different value. There are also data sets which can only be described by a specific measurement of central tendency. It is important that you understand which measurement is appropriate and which, if any, is best to describe the middle of your particular data set. (The **mode** is usually mentioned along with the mean and the median but it is not strictly a measure of the centre of a data set and so is discussed in the separate study guide: *The Mode*.)

The mean

The **mean** of a data set is what most people quote when they give an average. The correct term for this measurement of central tendency is the **arithmetic mean**; however it is so common that it is usually referred to simply as **the mean**. The mean is calculated by adding up all the values in a data set and then dividing by the number of pieces of data in that set.

Mathematically the calculation for the mean is described by the equation:

$$\bar{x} = \frac{\sum x}{n}$$

Where $\bar{x}$ (pronounced *x*-bar) is the mathematical symbol for the mean. Here each piece of data is described by x and the Greek capital letter sigma Σ, a very common symbol in statistics, tells you to add them all up. Finally, the number of pieces of data is given by n.

Example: Six students were asked about the amount of student loan they had taken out for a specific academic year. They said £3400, £2800, £4500, £3500, £3750 and £4625 respectively. What is the average loan for these students?

To use the formula above you need to know the x's which are the individual amounts of students loans (£3400, £2800, £4500, £3500, £3750, £4625) and n which is the number of students ($n = 6$). You can now use the formula:

$$\bar{x} = \frac{\sum x}{n} = \frac{3400 + 2800 + 4500 + 3500 + 3750 + 4625}{6} = \frac{22575}{6} = 3762.50$$

So the average (mean) loan is £3762.50.

The example illustrates that the calculation of the mean takes into account every value in a data set. This can be a disadvantage if your data are spread apart as the mean begins to lose its effectiveness as the mean is sensitive to individual data points which are very large (or very small) with respect to the rest of the data. These types of data are called **outliers** and the mean can be distorted by them.

Example: Calculate the mean wages for staff working in offices A and B:

Office A Salaries (thousand pounds): 15 18 17 14 16 18 15 17 12 20

Office B Salaries (thousand pounds): 15 18 16 14 15 16 12 17 90 95

Using the formula for the mean, for Office A (indicated by a subscript A in the calculation):

$$\bar{x}_A = \frac{\sum x}{n} = \frac{15+18+17+14+16+18+15+17+12+20}{10} = 16.2$$

So the mean wage for office A is £16 200. Similarly for Office B:

$$\bar{x}_B = \frac{\sum x}{n} = \frac{15+18+16+14+15+12+17+17+90+95}{10} = 30.8$$

So the mean wage for Office B is £30 800 even though most workers have a salary of around £16 000. These two results show that the mean wage is much higher in Office B. The mean of office B is distorted by the two high wages (outliers) of £90 000 and £95 000 and so it may not be the most suitable statistic to describe the centre of that data set. The middle of a data set with outliers is best described by the median (see following section). However as the spread of salaries for the workers in office A is much narrower, the mean is a good measure of the centre of this data set.

The mean is also unsuitable for non-interval levels of data. It makes no sense to use the method above to try and calculate the mean of a nominal data set. Could you have a mean favourite fruit for example? For ordinal data you may think that the mean is suitable (take questionnaires which ask you to rate your happiness on a scale of 1 to 10 for example) but it is not. Even though you can add up all the scores and then divide by the total number of correspondents each happiness score is subjective (a score of 8 does not necessarily mean twice as happy as a score of 4) and so the value of the mean loses its effectiveness.

The median

When data in a set is arranged (i.e. sorted) in ascending or descending order, the middle value of that list is described as the **median**. To find the median, you must consider the number of pieces of data in your data set. If the number is odd then the median is located at the exact centre of the list. If the number is even then the median is the mean value of the two central values. As you order the data, the median is not adversely affected by outliers and is the most suitable way for describing the centre of ordinal data.

Example: 11 families were asked how many children they had; the replies were 3, 4, 2, 5, 0, 0, 1, 3, 1, 1 and 2. What is the median number of children per family?

Firstly arrange the numbers in order and then pick the central value as you have an odd number of data.

0 0 1 1 1 **2** 2 3 3 4 5

The central value (in bold) is the 6th value as there are 5 values above and 5 values below it and so the median is 2.

Example: 10 students take a test and the results are as follows 18, 14, 7, 13, 12, 13, 12, 19, 10 and 11. What is the median score?

Again arrange the scores in order. Here you have an even number of results so you must find the middle two data which are the 6th and 7th.

7	10	11	12	**12**	**13**	13	14	18	19

taking the mean of these two values $(12+13)\div 2=12.5$ and so a score of 12.5 is the median.

The median is an excellent representation of central tendency when your data set contains outliers. If you look again at the salaries for Office A and Office B, arranged on order:

Office A:	12	14	15	15	**16**	**17**	17	18	18	20
Office B:	12	14	15	15	**16**	**16**	17	18	90	95

The median for Office A is £16 500 and for Office B is £16 000. Remember that the mean for Office A was £16 200 and the mean for Office B was £30 800. When you compare the mean and the median wages for Office A and Office B it becomes obvious just how much the outliers have affected the mean for Office B. Here it would be more sensible to report the median wage for Office B as the measure of central tendency.

Want to know more?

If you have any further questions about this topic you can make an appointment to see a **Learning Enhancement Tutor** in the **Student Support Service**, as well as speaking to your lecturer or adviser.

Call: 01603 592761

Ask: ask.let@uea.ac.uk

Click: https://portal.uea.ac.uk/student-support-service/learning-enhancement

There are many other resources to help you with your studies on our website.
For this topic there is a webcast.

Your comments or suggestions about our resources are very welcome.

Scan the QR-code with a smartphone app for a webcast of this study guide.